Washington, D.C. Neighborhood Analysis - Data Science Challenge

DSCI 551 Spring 2021
Madeleine Thompson

Background

I will be moving to Washington, D.C. over the summer for an internship and in about a year, after I graduate with my MS Applied Data Science, to accept a full-time position in the city. I am trying to collect data on different neighborhoods in and around Washington, D.C. so that I can make the most informed decision as to where I want to live. I want to be able to afford to buy a house in the neighborhood that I move to, I want to be close to work, and I want to have access to vegetarian and vegan restaurants. I've used the Yelp Fusion API to gather information on restaurants in each neighborhood (*restaurant_data.json*), the Google Maps Distance Matrix API to gather commute information (*commute_info.csv*), and I have scraped several web pages to populate a list of zip codes that I want to search over (*zip_codes.csv*).

Datasets

zip_codes.csv - 141 KB, columns=[state, zip_code, zip_url, latitude, longitude, population, house_value, household_income]
restaurant_data.json - 146 KB, structure={ zip_code: [RESTAURANT LIST { restaurant_id, ZIP_code, restaurant_name, address, city }] }
commute_info.csv - 64 KB, columns=[zip_code, latitude, longitude, distance, duration]

Data Science Challenges

Data Cleaning

1. Filter data based off of a commute of less than or equal to 1 hour driving time to work every day regardless of actual miles traveled. The work location is already specified and the commute information was queried based off of this pre-specified location.
2. Get rid of the data that does not meet this criteria to save space and unnecessary analysis of irrelevant data.

Data Integration

1. Merge all datasets together into a single file once filtered and cleaned.
    a. CSV or JSON

Data Analysis

1. Determine which neighborhoods would be good for me to live in. This is an optimization problem between these three "wants":
    a. Home prices are feasible on combined income.
        i. Between $200,000 and $400,000 USD
    b. Commute time as short as possible.
        i. Less than one hour to work
    c. Neighborhood has vegan/vegetarian restaurants or they are close by.